

Stat-Q User Manual

1. Overview

STAT-Q was first designed to calculate statistical comparison between simulated and observed hourly discharge time series. It can be used to conduct statistical analysis between any two time series of the same type **as long as the time series have the same time step**. The allowable time steps are 1, 2, 3, 4, 6, 8, 12, and 24 hours.

STAT-Q delivered as part of OHD-CORE-CHPS uses an approach similar to existing legacy models in OHD-CORE-CHPS where data provided by FEWS (inputs.xml, runInfo.xml, and params.xml) is formatted into text files for executing the existing standalone STAT-Q program. One difference is the output is not an outputs.xml time series file, but instead an output text file that is formatted for display through a web browser.

STAT-Q outputs:

- a summary of multi-year statistics,
- yearly statistics, and
- monthly statistics
- a table of 25 largest differences between the two data series.

And, it calculates statistics for a given number of evenly divided flows.

2. Utility Parameters

STAT-Q uses an XML representation of parameters where each parameter is captured within a separate XML tag. Each tag represents a pre-defined mapping of the original STAT-Q control file to a FEWS xml parameter element. Each parameter tag is internally transformed by the STAT-Q CHPS module into the required control file required by STAT-Q legacy utility.

Note: The parameter tag OPERATION_CONTENT is optional and if defined in the parameters file this parameter is expected to contain the contents of a legacy STAT-Q control file and it will be used as is.

All parameter tag and their arguments are converted into a single control file. The parameter and its value is translated to start with the symbol @ followed by a specific letter (except for file name and directory entry).

The following is an example of STAT-Q control file:

```
#This is the control file for stat_q program
#Do not comment immediately after a command, if it takes numerical arguments
#Type '@' & begin by entering the file names on the following line

@
6 1993 5 1999
1233

# the above line defines analysis period as start-month, start-year
# end-month, and end-year. Use negative values to use default period
# the fifth line defines the basin area in km^2, use a negative value
# if rainfall data were to be analyzed

# time step for analysis
@T 1

#display lines to std. output
@A 9

# yearly and monthly statistics are default output

#For a number of equal flow intervals
@D 5

#Values for flow_greater than given number of values
# first entry is the number of values
@E 2 250.0 350.0

#Values for flow intervals 4-5,1-2. 2 iterations.
@F 2 35.0 50.0 100.0 200.0

#Values for events. start time end time in hh mm dd yyyy format
@g 10
21 11 13 1994
6 11 17 1994

23 5 5 1995
11 5 11 1995

7 11 7 1996
8 11 10 1996

22 11 6 1996
8 11 10 1996
```

The table below shows the available parameter tags and their relation to the control file

Name	Type	Required [Yes/No]	Comment
START_DATE_MONTH_YEAR	String	No	@ The analysis period start year. The analysis period can only be specified for

			files with a complete month or more of data.
END_DATE_MONTH_YEAR	String	No	@ The analysis period end year
BASIN_AREA_IN_KM2	Double	Yes	@ a basin area (in km ²) has to be given if discharge time series are analyzed.
PROCESSING_TIMESTEP_IN_HOURS	Integer	Yes	@ Option to process data files for a specified time step other than what is specified in the time series data files
NUMBER_OF_LINES_OF_DATA_TO_DISPLAY	Integer	No	@A – Option for displaying the first specified lines of data.
NUMBER_OF_EQUAL_FLOW_INTERVALS	Integer	No	@D – Option for producing statistics in equal flow intervals.
CALCULATE_STATS_FOR_CASES_GREATER_THAN	Table	No	@E – Option for calculating statistics for cases that are greater than specified values.
PAIR_OF_FLOW_INTERVALS_IN_CMS	Table	No	@F – Option for calculating statistics of

			given flow intervals.
EVENT_START_AND_END_DATES_IN_MMDDYYYYHH	Table	No	@G -- Option for calculating statistics for events.
THRESHOLD_VALUE	Double	No	@H – Option to calculate statistics for events found above a given threshold value within a given time window.
THRESHOLD_WINDOW_LENGTH_HOURS	Integer	No	@H Option
PARTITION_NUMBER_MONTHS_FOR_ACCUMULATIVE_FLOW	Integer	No	@I – Operation to compute accumulated flow (in millimeters) for entire time period covered by data files.
OPERATION_CONTENTS	String	No	Contents of a legacy STAT-Q Control File If present all other parameters values will be ignored.

Example:

Module Parameter File:

[ModuleParFiles\STATQ_WLWA2_UpdateStates.xml](#)

3. Utility Time Series

STAT-Q model has 2 required input time series one observed and one simulated.

INPUT TIMESERIES:

Time Series Type	Internal Model Units	Time Step	Missing Values Allowed	Required [Yes or No]
Simulated		any	No	Yes
Observed		any	No	Yes

4. Utility Output

As mentioned STAT-Q output is not an outputs.xml time series file, but instead an output text file that is formatted for display through a web browser.

The following are examples of STAT-Q outputs:

- a summary of multi-year statistics,

MULTI-YEAR STATISTICS												
% Bias	Abs. % Bias	Obs. Qmean CMS	Sim. Qmean CMS	Obs. std CMS	Sim. std CMS	Obs. Cv	Sim. Cv	% RMS	RMS (CMS)	R	Nash-S. r	Modi. Rm
-3.31	32.0	3.30	3.19	3.71	3.92	1.13	1.23	57.5	1.90	0.878	0.739	0.832
Best line fit: Qobs = A+B*Qsim: A--> 0.646 (CMS) B--> 0.832												

- Yearly statistics

YEARLY STATISTICS									
Year	Percent Bias	Absolute Percent Bias	Absolute Error (CMS)	Observed Qmean	Simulated Qmean	Percent RMS	RMS (CMS)	R	Nash-S. r
1975	55.7	59.3	0.845	1.43	2.22	89.7	1.28	0.849	-1.33
1976	32.2	55.8	1.51	2.71	3.58	117.	3.18	0.760	0.325
1977	44.7	63.7	0.967	1.52	2.20	109.	1.66	0.705	-0.927
1978	26.1	45.8	1.39	3.03	3.81	73.0	2.21	0.923	0.360
1979	15.7	35.9	1.25	3.49	4.04	73.2	2.55	0.946	0.731
1980	17.7	36.9	0.952	2.58	3.04	60.3	1.56	0.876	0.246
1981	14.7	37.5	1.06	2.83	3.24	63.0	1.78	0.856	0.423
1982	7.48	31.5	1.10	3.47	3.73	54.8	1.90	0.905	0.777
1983	-5.75	33.2	1.14	3.44	3.24	49.8	1.72	0.889	0.788
1984	7.37	21.5	0.782	3.64	3.91	31.0	1.13	0.956	0.837
1985	17.0	27.7	0.961	3.47	4.06	48.5	1.68	0.954	0.763
1986	-21.0	28.0	1.34	4.80	3.79	44.8	2.15	0.921	0.794
1987	-15.2	32.1	0.993	3.09	2.62	46.0	1.42	0.827	0.580
1988	10.9	41.0	0.899	2.19	2.43	88.1	1.93	0.701	0.426
1989	-14.8	31.0	0.653	2.11	1.80	41.7	0.879	0.859	0.656
1990	-30.1	31.6	1.21	3.83	2.68	54.4	2.08	0.925	0.717
1991	0.352	27.4	0.993	3.63	3.64	37.8	1.37	0.902	0.796
1992	-22.4	31.3	0.927	2.96	2.30	40.5	1.20	0.845	0.552
1993	-15.2	29.1	1.48	5.08	4.31	39.3	2.00	0.937	0.856
1994	-20.8	30.5	0.745	2.44	1.93	41.3	1.01	0.936	0.833
1995	4.89	32.7	0.811	2.48	2.60	57.8	1.43	0.757	0.242
1996	-10.2	29.4	0.965	3.29	2.95	39.5	1.30	0.918	0.824
1997	-15.7	39.5	1.46	3.70	3.12	82.8	3.06	0.815	0.482
1998	-9.92	29.2	1.04	3.54	3.19	63.5	2.25	0.800	0.623
1999	8.46	27.2	0.992	3.66	3.96	43.0	1.57	0.930	0.830
2000	-9.81	25.6	1.10	4.30	3.88	42.8	1.84	0.931	0.859
2001	4.20	32.7	1.35	4.12	4.29	48.8	2.01	0.870	0.589
2002	-7.42	29.4	0.847	2.88	2.67	40.9	1.18	0.887	0.742
2003	-6.83	24.3	0.440	1.81	1.69	39.5	0.716	0.919	0.823
2004	6.11	22.9	0.816	3.57	3.79	59.8	2.14	0.944	0.813
2005	-19.7	37.9	0.762	2.01	1.62	65.3	1.31	0.750	0.484
2006	-25.2	28.1	1.00	3.56	2.66	43.6	1.55	0.904	0.713
2007	-15.2	25.5	0.943	3.70	3.14	38.7	1.43	0.936	0.852
2008	-14.2	31.5	2.35	7.45	6.39	56.3	4.19	0.902	0.801

- Monthly statistics

MONTHLY STATISTICS									
Month	Percent Bias	Absolute Percent Bias	Absolute Error (CMS)	Observed Qmean	Simulated Qmean	Percent RMS	RMS (CMS)	R	Nash-S. r
1	-9.32	35.7	0.762	2.13	1.93	60.2	1.28	0.758	0.496
2	2.45	32.8	1.01	3.08	3.16	48.8	1.50	0.886	0.730
3	2.11	30.7	1.68	5.48	5.60	50.5	2.77	0.817	0.570
4	2.48	28.9	1.83	6.34	6.49	40.7	2.58	0.878	0.725
5	-5.42	25.8	1.11	4.32	4.09	45.5	1.96	0.907	0.795
6	-2.45	29.1	1.11	3.81	3.72	72.8	2.77	0.891	0.790
7	-3.46	30.0	0.747	2.49	2.40	57.2	1.42	0.866	0.734
8	-4.16	40.2	0.913	2.27	2.18	86.4	1.96	0.749	0.521
9	-11.2	38.4	0.892	2.33	2.07	69.9	1.63	0.821	0.664
10	-17.2	36.5	0.828	2.27	1.88	55.0	1.25	0.861	0.633
11	-7.71	37.1	0.955	2.57	2.37	52.0	1.34	0.909	0.488
12	-3.42	33.1	0.840	2.54	2.45	48.3	1.23	0.866	0.615

- A table of 25 largest differences between the two data series.

25 LARGEST FLOW DIFFERENCES				
Date	Observed Flow CMS	Simulated Flow CMS	Error (Sim-Obs)	Percent Error
00:00 05/22/2004	29.183	48.592	19.409	66.5
00:00 03/06/1976	25.835	6.611	-19.224	-74.4
00:00 03/24/1979	19.928	39.070	19.142	96.0
00:00 06/12/2008	44.291	25.474	-18.817	-42.4
00:00 03/23/1979	22.826	40.030	17.204	75.3
00:00 03/07/1976	22.045	6.004	-16.040	-72.7
00:00 06/22/1997	35.181	51.152	15.971	45.3
00:00 01/31/1988	19.521	4.483	-15.038	-77.0
00:00 08/09/1998	22.140	7.998	-14.143	-63.8
00:00 03/04/1976	20.360	6.297	-14.063	-69.0
00:00 05/11/1990	31.638	17.790	-13.848	-43.7
00:00 05/12/1990	28.960	15.979	-12.981	-44.8
00:00 06/23/1997	33.378	20.679	-12.699	-38.0
00:00 04/04/1982	26.200	13.556	-12.643	-48.2
00:00 06/24/1997	26.868	14.244	-12.625	-46.9
00:00 06/15/2008	39.354	26.841	-12.513	-31.7

5. Configuring STAT-Q in FEWS

Configuring of the STATQ module assumes you have already set up CHPS Calibration. If you have not yet set up CHPS calibration configurations, do those first before configuring STAT-Q. Information about how to configure STAT-Q and Calibration can be found in "Installing the STATQ module".

6. FEWS Adapter Used

STAT-Q model uses the OHDFewsadapter to communicate with FEWS. Information about this adapter can be found at [OHDFewsadapter](#).

7. Common Issues

Issues	Possible causes
WARNING MESSAGE	
StartDate must be the first of the month, not "XX" moving date to first of next month.	The start date property is set different to the first of the month. STAT-Q only works with monthly set of dates.
StartDate "yyyy-MM-dd HH:mm:ss" must be the first day of the month, not "XX" moving date to first of next month."	STAT-Q only works with monthly sets of dates. That is the logic used to move the startDate to the beginning of the next month.

For the parameter "EVENT_START_AND_END_DATES_IN_MMDD YYYYHH" the "EVENT_END_DATES" date "XXX- XX-XX XX:XX:XX" will be removed as it is outside the analysis period. From XXX-XX-XX XX:XX:XX + " to: XXX-XX-XX XX:XX:XX	The date XXX-XX-XX XX:XX:XX defined in the parameter "EVENT_START_AND_END_DATES_IN_MMDDY YYYYHH" is not between the analysis period date defined in the control file or properties.
ERROR MESSAGE:	
The number of time series must be 2. One observed and one Simulated time series.	The input time series file have more than 2 time series defined or less than 2 time series defined. STAT-Q must have only 2 time series one observed time series and one simulated time series.
It must be 2 timeseries. One observed and one Simulated time series	One of the two time series in the input file must be of type simulated and other of type observed.
The simulated "simulatedParameterId" and observed "observedParameterId" time series must have same units. Please check the workflow property at XXXXX_Stats_calibration.xml"	The unit type defined for the simulated and observed time series must be the same.
Discharge data must have units of CMS	If discharge data is present. The time series must have units of CMS
Rainfall data must have units in MM	If rainfall data is present. The time series must have units of MM
The parameter "PROCESSING_TIMESTEP_IN_HOURS" must be divisible by the timeseries interval in hours	
The parameter "PROCESSING_TIMESTEP_IN_HOURS" must be greater or equal to the timeseries interval in hours	The parameter PROCESSING_TIMESTEP_IN_HOURS is smaller than expected. It must be greater or equal to the observed or simulated time series interval.
The start date from the control file "START_DATE_MONTH_YEAR" "mmyyyy" must be greater or equal to the timeseries start Date time "MMDDYYYYHH" Check parameter file in ModuleParFiles directory	The START_DATE_MONTH_YEAR parameters is smaller that expected.
The end date from the control file " END_DATE_MONTH_YEAR " "XXXX" must be less or equal to the timeseries End Date time "YYYYY"	The value of the parameter END_DATE_MONTH_YEAR is bigger than the end date of the simulated or observed time series.
The date Parameter "START_DATE_MONTH_YEAR " is bigger than date parameter "END_DATE_MONTH_YEAR"	The analysis period defined in the parameter file is wrong. Start date must be minor that end date.
One of the values of the "PAIR_OF_FLOW_INTERVALS_IN_CMS" parameter is missing. It must have exactly 2 columns per row"	Only one column was defined for the parameter "PAIR_OF_FLOW_INTERVALS_IN_CMS" but two are required.

The value "XXXX" of the "EVENT_START_AND_END_DATES_IN_MMDDYYYYHH" parameter is wrong. It must have a valid date using format "MMDDYYYYHH"	The date format of the parameter EVENT_START_AND_END_DATES_IN_MMDDYY YYHH is wrong. It could have an invalid or missing value.
The Parameter "PARTITION_NUMBER_MONTHS_FOR_ACCUMULATIVE_FLOW" must be greater or equal than 1 and less or equal than 12	The parameter value must be between 1 and 12.